UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/788,455	03/01/2004	Kent Bodell	C697 0007/GNM	7385

720 7590 10/07/2009

OYEN, WIGGS, GREEN & MUTALA LLP
480 - THE STATION
601 WEST CORDOVA STREET
VANCOUVER, BC V6B 1G1
CANADA

EXAMINER	
GUPTA, MUKTESH G	

ART UNIT	PAPER NUMBER
2444	

NOTIFICATION DATE	DELIVERY MODE
10/07/2009	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

mail@patentable.com

Office Action Summary	Application No.	Applicant(s)	
	10/788,455	BODELL ET AL.	
	Examiner	Art Unit	
	Muktesh G. Gupta	2444	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *20 July 2009*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-20* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. **Claims 1-5, 7, 11-12 and 15-20** are amended.

   Claims 1-20 are presented have been examined on merits and are pending in this application.

Continued Examination Under 37 CFR 1.114

2. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 07/20/2009 has been entered.

Response to Amendments/Arguments

3. Acknowledgment is made for Applicants Amendments for claims filed on 07/20/2009.

   Applicant's arguments are deemed moot in view of the following new grounds of rejection as explained here below, necessitated by Applicants substantial amendments to Claims.

   Applicant's arguments with respect to **Claims 1-20,** have been considered but are moot in view of the new ground(s) of rejection.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

4. **Claims 1-5, 8-12 and 15-20** rejected under 35 U.S.C. 102(e) as being anticipated by US Patent Publication No. 20030208531 to Matters, Todd et al., (hereinafter "Matters").

As to Claims 1, 3, 11, 17-18 and 20, *Matters discloses method for communicating data from a first compute node of a computer system to a second node of the computer system, the computer system comprising multiple compute nodes, including the first and second compute nodes, interconnected by an inter-node communication network, the method comprising* (as stated in par. [0013-0014], Matters discloses a computer system that includes a plurality of servers, and a shared I/O subsystem coupled to each of the servers and to one or more I/O interfaces. The servers are interconnected to the shared I/O subsystem by a high-speed, high-bandwidth, low-latency switching fabric. The switching fabric includes dedicated circuits, which allow the various servers to communicate with each other. In one embodiment, the switching fabric uses the InfiniBand protocol for communication):

placing the data on a full-duplex packetized interconnect directly connecting a CPU of the first compute node to a first network interface of the first compute node, the first network interface directly connected to the inter-node communication network (as stated in par. [0017], par. [0050], par. [0068], par. [0075], par. [0083], Matters discloses shared I/O subsystem that couples a plurality of computer systems to at least one shared I/O interface. The shared I/O subsystem includes a plurality of virtual I/O interfaces that are communicatively coupled to the computer systems where each of the computer systems includes a virtual adapter that communicates with one of the virtual I/O interfaces. The shared I/O subsystem further includes a forwarding function having a forwarding table that includes a plurality of entries corresponding to each of the virtual I/O interfaces. The forwarding function receives a first I/O packet from one of the virtual I/O interfaces and uses the forwarding table to direct the first I/O packet to at least one of a physical adapter associated with the at least one shared I/O interface and one or more of other ones of the virtual I/O interfaces. Also, using shared I/O subsystem 60, a server 5 can connect directly to existing network sources such as network storage 85 or even the Internet 80 via respective I/O interface units 62. I/O interface units 62 are used as line cards to provide a connection to multiple computer systems. I/O interface unit 62 may also be connected to an existing network system, such as an Ethernet or other types of network system. Thus, I/O interface unit 62 can include a Target Channel Adapter (TCA) 217 for coupling network links. Internal Protocol is a protocol used in shared I/O subsystem 60 to supports full duplex packet passing within I/O management link 236. Using Internal Protocol over I/O link layer 274 (shown in FIG. 5C), shared I/O

subsystem 60 can support various protocols between each I/O interface unit 62 and between I/O interface units 62 and switch card 228. Driver 272 is a software device driver on the main CPU (not shown) of I/O interface unit 62/switch card 228. Driver 272 will be fully responsible for the physical interface between the main CPU (not shown) of I/O interface unit 62/switch card 230 and its IBML interface hardware. This interface may be a high speed serial port on a CPU or other interfaces. FIG. 6 shows one embodiment of shared I/O subsystem 60 using I/O interface unit 62 coupled to multiple servers 255. The embodiment as shown has I/O interface unit 62 configured for use with InfiniBand protocols such as IBML protocol. On each server 255, HCA 215 performs all the functions required to send/receive complete I/O requests. HCA 215 communicates to I/O interface unit 62 by sending I/O requests through a fabric, such as InfiniBand fabric 160 shown in the diagram. As it is apparent from the diagram, typical network components such as NIC 40 and HBA 50 (shown in FIG. 1A) have been replaced with HCA 215);

receiving the data at the first network interface; and, transmitting the data from the first network interface to a second network interface of the second compute node by way of the inter-node communication network (as stated in par. [0093], Matters discloses an outbound packet (of data) originates in protocol stack 221 and is delivered to virtual NIC 222. Virtual NIC 222 encapsulates the packet into a combination of Send/Receive and Remote Direct Memory Access (RDMA) based operations which are delivered to HCA 215. These Send/Receive and RDMA based operations logically form virtual I/O bus 240 interface between virtual NIC 222 and virtual port 242. The

operations (i.e., packet transfers) are communicated by HCA 215, through InfiniBand links 165 and InfiniBand fabric 160 to TCA 217. These operations are reassembled into a packet in virtual port 242. Virtual port 242 delivers the packet to switching function 250. Based on the destination address of the packet, forwarding table 245 is used to determine whether the packet will be delivered to another virtual port 242 or NIC 40, which is coupled to network systems 105).

As to Claims 2 and 12, *Matters discloses method according to claims 1 and 11, wherein the first network interface and the CPU are the only devices configured to place data on the packetized interconnect* (as stated in par. [0059], par. [0088], Matters discloses as illustrated, in FIG. 2C, only steps 502 and 504 take place at a server or host level. All other steps take place at the shared I/O subsystem level. In step 502, applications from one or more hosts (e.g., server) form I/O requests. Typical I/O requests may include any programs or operations that are being transferred to the dedicated I/O subsystem. In step 504, multiple I/O requests from multiple hosts are sent to shared I/O subsystem 60. Using virtual NIC 222, server 255 communicates via virtual I/O bus 240, which connects to virtual port 242. Virtual port 242 exists within I/O interface unit 62 and cooperates with virtual NIC 222 to perform typical functions of physical NICs 40. Note that virtual NIC 222 effectively replaces the local PCI bus system 20 (shown in FIG. 1A), thereby reducing the complexity of a traditional server system. In accordance with one aspect of the present invention, a physical NIC 40 is "split" into multiple virtual NICs 222. That is, only one physical NIC 40 is placed in I/O

interface unit 62. This physical NIC 40 is divided into multiple virtual NICs 222, thereby allowing all servers 255 to communicate with existing external networks via I/O interface unit 62. Single NIC 40 appears to multiple servers 255 as if each server 255 had its own NIC 40. In other words, each server "thinks" it has its own dedicated NIC 40 as a result of the virtual NICs 222).

As to Claims 4 and 16, *Matters discloses method according to claims 3 and 11, comprising passing the data through a buffer at the first network interface before transmitting the data* (as stated in par. [0098-0099], Matters discloses One embodiment of the present invention uses virtual port frame 380 (shown in FIG. 7B) to exchange data between each virtual port 242 and between virtual port 242 and a physical I/O interface such as NIC 40, all of which are shown in FIG. 7A. A virtual port 242 arranges (or writes) data into virtual port frame 380 (shown in FIG. 7B). Upon completion of write, virtual port frame 380 is transmitted to a buffer in shared I/O subsystem 60. Shared I/O subsystem 60, by detecting control bits contained in virtual port frame 380, recognizes when the transmission of data is completed. Thereafter, shared I/O subsystem 60 forwards the data packet to an appropriate virtual port 242).

As to Claims 5 and 19, *Matters discloses method according to claims 1 and 11, comprising, at the first network interface, determining a size of the data and, based upon the size of the data, selecting among two or more protocols for transmitting the data* (as stated in par. [0100], par. [0106], Matters discloses embodiment of using the

Internal Protocol described above can be used to exchange data that follows many different protocols. For instance, virtual ports 242 can exchange virtual port frames 380 to communicate Ethernet frame data. That is, the virtual port frames 380 can be used to send/receive Ethernet data having a variable length among virtual ports 242 and NIC 40 without using interrupt signals. Virtual port frame 380 can be used to transfer data that follows various protocols, and as such, using other data that follow different protocols (and variable length) is within the scope of the present invention).

As to Claims 8, 10 and 15, *Matters discloses method according to claims 1 and 11, wherein the data comprises a raw ethertype datagram and transmitting the data comprises encapsulating the raw ethertype datagram within one or more link layer packet headers* (as stated in par. [0081], par. [0090-0091], par. [0093], Matters discloses Link layer 274 implements the Internal Protocol, and provides for fragmentation and reassembly of data frames. Link layer 274 expects in order delivery of packets and provides an unreliable datagram link layer. To the layers above it, an Ethernet API will be presented. In accordance with one aspect of the present invention, as shown further in FIG. 7A, all I/O requests and other data transfers are handled by HCA 215 and TCA 217. As noted above, within each server 255, there are multiple layers of protocol stacked on the top of HCA 215. As shown, virtual NIC 222 sits on top of HCA 215. On top of virtual NIC 222, a collection of protocol stack 221 exists. Protocol stack 221, as shown in FIG. 7A, includes link layer driver 223, network layer 224, transport layer 225, and applications 226. Virtual NIC 222 exists on top of HCA 215.

Link layer driver 223 controls the HCA 215 and causes data packets to traverse the physical link such as InfiniBand links 165. Above link layer driver 223, network layer 224 exists. Network layer 224 typically performs higher level network functions such as routing. For instance, in one embodiment of the present invention, the network layer 224 includes popular protocols such as Internet Protocol (IP) and Internetwork Packet Exchange.TM. (IPX). Above network layer 224, transport layer 225 exists. Transport layer 225 performs even higher level functions, such as packet assembly/fragmentation, packet reordering, and recovery from lost or corrupted packets. In one embodiment of the present invention, the transport layer 225 includes Transport (or Transmission) Control Protocol (TCP). An outbound packet (of data) originates in protocol stack 221 and is delivered to virtual NIC 222. Virtual NIC 222 encapsulates the packet into a combination of Send/Receive and Remote Direct Memory Access (RDMA) based operations which are delivered to HCA 215).

As to Claim 9, *Matters discloses method according to claim 8 wherein the link layer packet headers comprise InfiniBand.TM. Link layer packet headers* (as stated in par. [0091], Matters discloses Virtual NIC 222 exists on top of HCA 215. Link layer driver 223 controls the HCA 215 and causes data packets to traverse the physical link such as InfiniBand links 165. Above link layer driver 223, network layer 224 exists. Network layer 224 typically performs higher level network functions such as routing).

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

5. **Claims 6-7 and 13-14** rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent Application Publication No. 20030208531 to Matters, Todd et al., (hereinafter "Matters") as applied to **Claims 6-7 and 13-14**, and in view of U.S. Patent No. 6542513 to Franke; Hubertus et al., (hereinafter "Franke").

Regarding **Claims 6-7 and 13-14,** Matters disclosed the invention substantially as claimed regarding using two or more protocols for transmitting the data. Matters do not explicitly disclose "two or more protocols comprise an eager protocol and a rendezvous protocol".

Matters discloses, as stated in par. [0081], par. [0100], par. [0106], Link layer 274 implements the Internal Protocol, and provides for fragmentation and reassembly of data frames. Link layer 274 expects in order delivery of packets and provides an unreliable datagram link layer. To the layers above it, an Ethernet API will be presented. Using the Internal Protocol described above can be used to exchange data that follows many different protocols. For instance, virtual ports 242 can exchange virtual port frames 380 to communicate Ethernet frame data. That is, the virtual port frames 380

can be used to send/receive Ethernet data having a variable length among virtual ports 242 and NIC 40 without using interrupt signals. As noted, virtual port frame 380 can be used to transfer data that follows various protocols, and as such, using other data that follow different protocols (and variable length) is within the scope of the present invention. Matters do not show protocols comprise an eager protocol and a rendezvous protocol.

Franke shows in a message processing system having message source and destination nodes, FIG. 5 is a protocol diagram of a second, eager rendezvous transmission mode in which message transmission is initiated using a packet having both control information and a data portion of the message, with any remaining data portions of the message being transmitted following an acknowledgement from the destination node, in an analogous art for the purpose of Matters (as suggested in title and abstract).

It would have been obvious to a person of ordinary skill in the art at the time of the invention was made to modify Matters teachings on managing server farms, to add the teachings of Franke.

The modifications would have been obvious because one of ordinary skill in the art would have been motivated for a method, system, and associated program code and data structures, protocols and buffering for facilitating the efficient transmission of messages from a source node to a destination node in a message processing system which prevent the performance degradation associated with packet retransmission after timeouts.

Together Matters and Franke disclosed all limitations of **Claims 6-7 and 13-14,** are rejected under 35 U.S.C. 103(a).

As to Claim 6, *Together Matters and Franke disclose method according to claim 5 wherein the two or more protocols comprise an eager protocol and a rendezvous protocol* (as stated by Frank in Abstract lines 1-12, An "eager" rendezvous transmission mode is disclosed in which early arrival buffering is provided at message destination nodes for a predetermined amount of data for each of a predetermined number of incoming messages. Relying on the presence of the early arrival buffering at a message destination node, a message source node can send a corresponding amount of message data to the destination node along with control information in an initial transmission).

As to Claim 7, *Together Matters and Franke disclose method according to claim 6 comprising, upon selecting the rendezvous protocol, automatically generating a Ready To Send message at the network interface of the first compute node* (as stated by Frank in Abstract lines 12-24, Any remaining message data is sent only upon receipt by the source node of an acknowledgement from the destination node indicating that the destination node is prepared to receive any remaining data. In an enhanced embodiment, the source node alternates between rendezvous transmission modes as a function of the amount of free space in the early arrival buffering at the destination node, as indicated by the number of outstanding initial transmissions for which

acknowledgements have not yet been received. Different transmission modes for different destination nodes can be employed at a source node, depending on the amount of early arrival buffering currently available in each respective destination node).

As to Claim 13, *Together Matters and Franke disclose compute node according to claim 11 comprising a memory and a facility configured to allocate eager protocol buffers in the memory and to automatically signal to one or more other compute nodes that the eager protocol buffers have been allocated* (as stated by Frank , in col. 6, lines 30-49, col. 11, lines 19-26, The length "N" of the first data portion of the message transmitted is a predetermined number which should correspond to the size "N" of the early arrival buffer slot pre-allocated at the destination node for each message of a number "Q" of messages. If "N" is large enough so that the destination node receives the control information in the initial transmission and returns the rendezvous acknowledgement before all "N" bytes have been sent by the source node, then the source node does not experience any interruption in the data transmission and the destination, likewise, does not see any interruption in the data received. This eager rendezvous transmission mode does not require a large amount of buffering at the destination for these initial transmissions since each such early arrival transmission only brings with it, at most, "N" bytes of the data portion of the message. The early arrival buffering may be a "flat" buffer, possibly from a buffer pool, or can be implemented using pointers and/or linked lists. The principles of transmission mode 300 of FIG. 6 can be applied across a system having multiple source and destination nodes. Each

source node maintains the count C of each of the unacknowledged initial eager rendezvous transmissions sent to each respective destination node, and alternates between the rendezvous transmission modes 100 and 200 on a per destination node basis, depending upon the count C for each respective destination node).

As to Claim 14, *Together Matters and Franke disclose compute node according to claim 13 comprising a facility configured to automatically associate memory protection keys with the eager protocol buffers and a facility configured to verify memory protection keys in incoming eager protocol messages before writing the incoming eager protocol messages to the eager protocol buffers.* (as stated by Matters, in par. 0046, lines 1-5, par. 0063, lines 1-5, par. 0064, lines 1-10, the distributed computer system shown in FIG. 1 performs operations that employ virtual addresses and virtual memory protection mechanisms to ensure correct and proper access to all memory. A RDMA Write work queue element provides a memory semantic operation to write a virtually contiguous memory space on a remote node. The RDMA Write work queue element contains a scatter list of local virtually contiguous memory spaces and the virtual address of the remote memory space into which the local memory spaces are written. A RDMA FetchOp work queue element provides a memory semantic operation to perform an atomic operation on a remote word. The RDMA FetchOp work queue element is a combined RDMA Read, Modify, and RDMA Write operation. The RDMA FetchOp work queue element can support several read-modify-write operations, such as Compare and Swap if equal. A bind (unbind) remote access key (R_Key) work queue element

provides a command to the host channel adapter hardware to modify (destroy) a memory window by associating (disassociating) the memory window to a memory region. The R_Key is part of each RDMA access and is used to validate that the remote process has permitted access to the buffer).

Remarks

6. The following pertaining arts are discovered and not used in this office action. Office reserves the right to use these arts in later actions.

a. Chui, Terence (US 20040034795 A1) Multi-service optical infiniband router

b. Foster, Michael S. et al. (US 20020181395 A1) Communicating data through a network so as to ensure quality of service

c. Johnsen; Bjorn Dag et al. (US 7443860 B2) Method and apparatus for source authentication in a communications network

d. Pekkala, Richard E. et al. (US 20020172195 A1) Apparatus and method for disparate fabric data and transaction buffering within infiniband device

e. Zhu; Julianne Jiang et al. (US 20080184008 A1) DELEGATING NETWORK PROCESSOR OPERATIONS TO STAR TOPOLOGY SERIAL BUS INTERFACES

Conclusion

7. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Muktesh G. Gupta whose telephone number is 571-270-

5011. The examiner can normally be reached on Monday-Friday, 8:00 a.m. -5:00 p.m., EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, William C. Vaughn can be reached on 571-272-3922. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

MG

/William C. Vaughn, Jr./

Supervisory Patent Examiner, Art Unit 2444